JAIST Repository
https://dspace.jaist.ac.jp/

| | |
|---|---|
| Title | 特許間引用のカテゴリ化 : 被引用数計測の精緻化に向けて |
| Author(s) | 鈴木, 裕; 芳鐘, 冬樹 |
| Citation | 年次学術大会講演要旨集, 27: 760-764 |
| Issue Date | 2012-10-27 |
| Type | Conference Paper |
| Text version | publisher |
| URL | http://hdl.handle.net/10119/11132 |
| Rights | 本著作物は研究・技術計画学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Science Policy and Research Management. |
| Description | 一般講演要旨 |

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

２Ｇ２８

# 特許間引用のカテゴリ化:被引用数計測の精緻化に向けて

○鈴木裕（筑波大学大学院），芳鐘冬樹（筑波大学）

## 1 はじめに

論文や特許といった膨大な数の文献情報の集合に対し、定量的な評価がなされることがしばしばあり、その評価尺度の一つとして、それがどれだけ他の文献内で引用されているかを表す数値である「被引用数」がある。この数値が高いことは、他の文献に利用されることが多いことを意味し、被引用数はある種の重要度を示すと考えられている。

特許の重要度と被引用数の関係については欧米の研究者らによって多くの研究が進められており、Narin は、アメリカの特許に関する訴訟において特許の技術的影響力を示すために引用が指標として利用されていると述べており [1]、Harhoff らは米国やドイツの特許では経済的価値の高いものがより多く引用されていることなどを実証している [2]。日本でも重要特許の判別指標として被引用数が有効であるとする後藤らの研究 [3]、引用情報に基づく特許文献の重要度算出方式を検討している佐藤らの研究 [4]などがある。

これらの研究の結果から、被引用数と特許の重要度には密接な関係があることが示唆され、被引用数が高ければ重要度も高いという解釈も成り立ちうる。しかし、安直に「被引用数＝重要度」とする見方が広がることを懸念する意見もあり、例えば三原は「引用行動パターンはさまざまであることから、それらの重みを一律に扱うのは極めて危険である」と指摘している [5]。そのような意見がありながら、特許の引用行動パターン（動機）についての研究はこれまでほとんど行われてこなかった。その理由は、特許文献の内容は出願する技術の新規性・進歩性を示すものであり、そこにある従来技術の引用は「新規性を示すためのもの」という一括りで認識されてきたからではないかと推測する。一方、論文も「新しい知見を示す」ことを目的とした文献であるが、論文引用の動機についての類型化は以前から行われている。例えば Garfield によって「先駆者への敬意」、「手法、技術の承認」といった 15 種類の動機が定義されている [6]。また、Teufel らは引用を 12 種類に分類した上で、それぞれの記述の特徴から自動的に分類が可能か機械学習を試みている [7]。

特許についても、どのように引用されているかを考慮することは重要である。例えば被引用数が同数の特許が 2 件あったとして、一方は技術の問題点の指摘ばかりであり、もう一方は引用側の特許と併用しうるものや特許技術の一部を構成するものとして記述されていたとすれば、後者の方がより技術的貢献が大きい（技術的価値がある）と考えられるからである。そのような引用の動機に重みを与える評価が可能な分析を実現する環境を整えるため、本研究では、特許が他の特許を引用する動機のカテゴリ化を試み、その分布状況について調査した。

## 2 データ

情報源として用いるのは、国立情報学研究所が提供する NTCIR-7[8]および NTCIR-8[9]のテストコレクションである。テストコレクションのうち、1998～2002 年（NTCIR-7）、2003～2007 年（NTCIR-8）の 10 年間に公開された日本国公開特許公報全文データ 3,546,429 件を本研究に使用する。

10 年間のデータから、各特許について、公開番号、出願番号、国際特許分類、被引用特許（その特許が引用している特許）の公開/出願番号を抽出した。また、1998 年公開の特許を標本特許とし、1998 年以降の特許が標本特許を引用している記述部分を抽出した。

標本特許の各技術分野（国際特許分類のセクション階層）において、引用年齢（特許が公開されてから引用されるまでの年数）の増加にともなう被引用数の推移を図 1 に示す。公開から 2 年後の 2000 年以降各分野とも大きく増加していき、2003 年に一度落ち込むものの、おおむね公開後 7 年（2005 年）まで増加傾向が続き、その後減少していく。2003 年の落ち込みは図 2 で示す通り公開数が前後の年に比して低いことによる影響が考えられる。

3 手法

本研究では、引用の動機が記述内容から比較的肯定的と判断できるものを「Type1」、比較的否定的と判断できるものを「Type2」、どちらともいえず判断できないものを「その他」とする 3 種類のカテゴリを設定した。

「Type1」に属する引用は、「引用特許内で有用な技術として説明されている」、「引用特許内で利用されている（利用可能である）」、「引用特許とともに用いられている」などである（例：特開２００７－３１６４４における「なお、本発明の原料となるアニリン化合物は特開平１０－２９９７６号公報に記載の方法を参考に合成することができる」）。「Type2」に属する引用は、「問題点・不足点が指摘されている」、「引用特許を含め、より上位の技術が挙げられている」、「前例がないことを示すための例」などである（例：特開２００７－８９８４５における「よって、下記特許文献１における安全引き具は、安全性の観点から難点を有するといわざるを得ない」）。「その他」に属する引用は、問題点または利点に言及していると判断し難いなど、Type1 とも Type2 とも分類できない引用である。

1998～2007 年公開の特許において標本特許を引用している箇所から、各年 200 件ずつ、合計 2,000 件をランダムに選び、それらの記述内容を精読して引用の分類を行う。引用を含む文とその前後 1 文ずつを合わせた引用箇所のコンコーダンス（文脈付き索引）を作成して、それを用いて分類を試みたが、半数以上の引用は、コンコーダンスのみではいずれのカテゴリに属すか判断できなかった。そのため、1 件 1 件、公報全文を参照しつつ分類を行うことにした。それぞれのカテゴリに属する引用の数の分布を調べ、また、その経年変化を追う。

4. 結果

まず、1998～2007 年の 10 年間を対象に分類した結果について、標本特許の技術分野別および全体（2,000 件）の、各カテゴリに属す引用の数を表 1、比率を表 2 にまとめた。なお、複数の技術分野に属する特許に関しては重複して計測しているため、各分野に属する引用の合計は 2,000 を超えている。

数で見た場合、各分野のばらつきは大きいものの、比率で見た場合はいずれも 9 割超の引用を Type1 または Type2 に分類することができた。Type1 と Type2 を比較すると、いずれも 6 割超が Type2 であり、7 割以上になる分野が多い。ここから、多くの場合、特許は比較的否定的に引用される傾向が読み取られる。2,000 件全体の比率を基準とすると、Type1 比率が高いものは C（化学；冶金）、次いで D（繊維；紙）、Type2 比率が高いものは E（固定造形物）、F（機械工学；照明；加熱；武器；爆破）、次いで H（電気）である。ただし、D（繊維；紙）および E（固定造形物）に属する特許の引用は特に少数であるため、誤差が大きい可能性がある。

さらに、Type1 比率の経年変化を示したのが図 3 である。誤差による変動の大きさに配慮して、D（繊維；紙）および E（固定造形物）は図から除いている。W（全体）に注目すると、標本特許の公開直後は 4 割超を占めるが、2 年で 2 割程度と半減し、以降はそれほど変化しない。分野別でも、やはり標本特許の公開直後が最も高く、2 年ほどで大きく減少し、そこからはあまり激しい変化はない。ただし、F（機械工学；照明；加熱；武器；爆破）については、公開直後の時点で既に Type1 比率が低いが、そこからさらに大きく減少するのは、他の分野より遅く 4 年後である。これらの結果は、(1) 特許の公開直後、技術内容が最新である時点では、比較的、肯定的な文脈で参照される傾向があるが、その後、技術の陳腐化や代替技術の発明が進むにともない否定的な参照の割合が増えていくこと、(2) その変化の速度は、技術分野によって異なること、(3) その後、肯定的な参照の割合の減少はおさまる、つまり、特許に有用性が見出されて参照される割合はある程度の水準で、しばらく保たれ続けることを示唆している。

5 おわりに

価値評価の指標の一つとされる被引用数の信頼性を高めるための重み付け要素として、特許の引用動機に着目し、そのカテゴリ化を試みた。動機を比較的肯定的、比較的否定的に二分するという粗いカテゴリ化であるが、多くの場合、特許は比較的否定的に引用されるという傾向や、引用年齢の増加にともない比較的肯定的な引用の割合が減少する傾向などが観察できた。

比較的否定的な引用より比較的肯定的な引用の方が高い有用性が認められていると考えれば、比較的肯定的な引用に重み付けをして特許の重要度を計測するのが妥当であろう。ただし、技術分野によって肯定的・否定的引用の比率の傾向に違いがあるため、分野横断的な重要度の比較においては注意を要する。

今後は、より細分化したカテゴリの設定や、記述の特徴に基づく機械学習による引用の自動分類を行いたい。

参考文献


1) Narin, F. Patents as indicators for the evaluation of industrial research output. Scientometrics. 1995, vol. 34, no. 3, p. 489-496.
2) Harhoff, D.; Narin, F.; Scherer, F. M.; Vopel, K. Citation frequency and the value of patented inventions. The Review of Economics and Statistics. 1999, vol. 81, no. 3, p. 511-515.
3) 後藤晃, 玄場公規, 鈴木潤, 玉田俊平太. “重要特許の判別指標”. RIETI-独立行政法人経済産業研究所. http://www.rieti.go.jp/jp/publications/summary/06030014.html, (参照 2012-09-16).
4) 佐藤祐介, 岩山真. 引用情報に基づく特許文献の重要度算出方式の検討. 情報処理学会研究報告. 2006, vol. 2006, no. 59, p. 9-16.
5) 三原健治. バイオテクノロジー分野における特許分類および引用情報を指標とした特許の価値評価に関する一考察. 情報管理. 2012, vol. 54, no. 11, p. 738-749.
6) Garfield, E. “Can citation indexing be automated?”, Essays of an Information Scientist. 1962, Vol. 1, p. 84-90.
7) Teufel, S.; Siddharthan, A.; Tidhar, D. Automatic classification of citation function. Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. 2006, p. 103-110.
8) Nanba, H.; Fujii, A.; Iwayama, M.; Hashimoto, T. Overview of the patent mining task at the NTCIR-7 workshop. In: Proceedings of NTCIR-7 workshop meeting (pp. 325-332). Tokyo: National Institute of Informatics. 2008.
9) Fujii, A.; Utiyama, M.; Yamamoto, M.; Utsuro, T.; Ehara, T.; Echizen-ya, H.; Shimohata, S. Overview of the patent translation task at the NTCIR-8 workshop. In: Proceedings of NTCIR-8 workshop meeting (pp. 371–376). Tokyo: National Institute of Informatics. 2010.

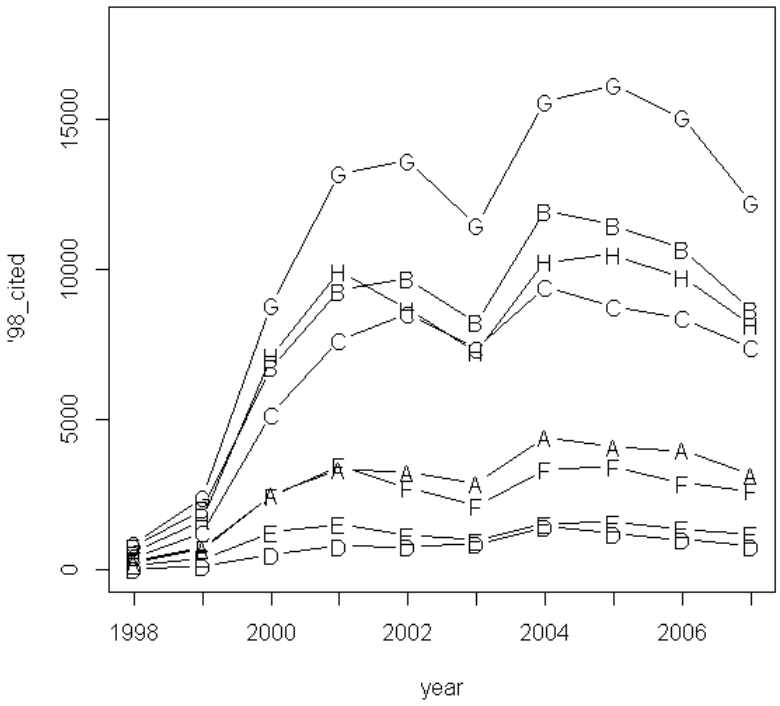




図 1：標本特許のセクション別に見た被引用数の推移

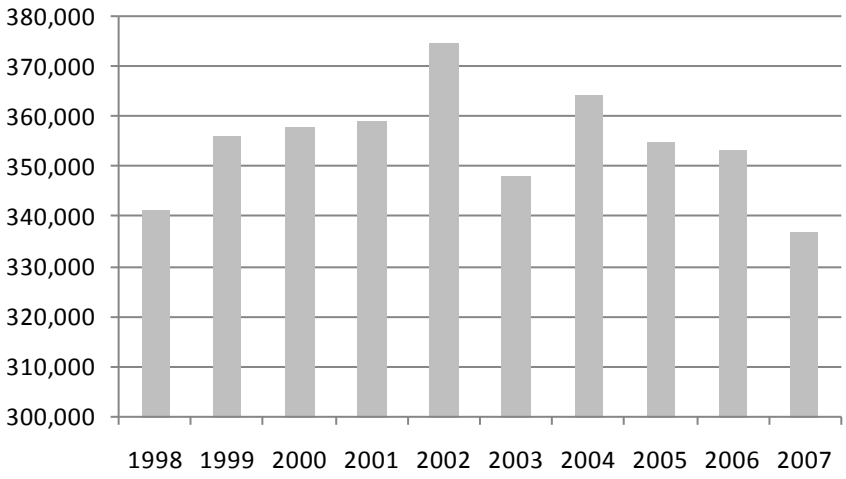



図 2：各年の特許公開件数

表 1：1998～2007 年全体のカテゴリ別引用数

| | Type1 | Type2 | その他 | 合計 |
|---|---|---|---|---|
| A 生活必需品 | 43 | 163 | 13 | 219 |
| B 処理操作；運輸 | 120 | 412 | 20 | 552 |
| C 化学；冶金 | 135 | 235 | 19 | 389 |
| D 繊維；紙 | 12 | 25 | 3 | 40 |
| E 固定造形物 | 8 | 68 | 4 | 80 |
| F 機械工学；照明；加熱；武器；爆破 | 19 | 161 | 9 | 189 |
| G 物理学 | 202 | 529 | 40 | 771 |
| H 電気 | 65 | 428 | 32 | 525 |
| 全体 | 415 | 1482 | 103 | 2000 |

表 2：1998～2007 年全体のカテゴリ別引用比率

| | Type1 | Type2 | その他 |
|---|---|---|---|
| A 生活必需品 | 20% | 74% | 6% |
| B 処理操作；運輸 | 22% | 75% | 4% |
| C 化学；冶金 | 35% | 60% | 5% |
| D 繊維；紙 | 30% | 63% | 8% |
| E 固定造形物 | 10% | 85% | 5% |
| F 機械工学；照明；加熱；武器；爆破 | 10% | 85% | 5% |
| G 物理学 | 26% | 69% | 5% |
| H 電気 | 12% | 82% | 6% |
| 全体 | 21% | 74% | 5% |

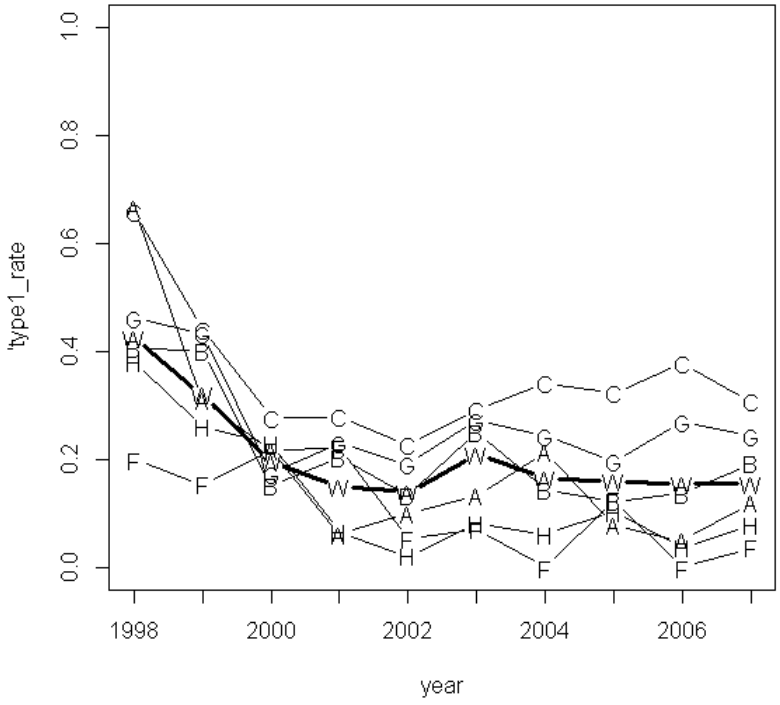



図 3：標本特許のセクション別に見た Type1 比率の推移（DE 除外）